APPENDIX 2. STATISTICAL METHODS FOR PRIMARY CPFV SURVEY

Vessel Directory Telephone / Intercept

The Vessel Directory Telephone Survey (VDTS) data was used to estimate the total numbers of angler fishing trips on charter boats included in the sample frame. Estimates of the numbers of boat and angler fishing trips were stratified by two-month wave and sampling week. In addition, angler trip data was post-stratified by primary area fished to directly estimate angler trips at the area level of stratification without using data from the Observer Survey. For each sampling week, the number of angler fishing trips on boats in the sample frame was independently estimated for each of the three primary areas of fishing. For the purpose of estimating total angler trips, we assumed simple random cluster sampling where each boat comprised a cluster of boat trips that were completely sampled. Point estimates of the numbers of boat trips and angler trips and the variances of those point estimates were calculated using the appropriate equations for estimation of a population total under simple random cluster sampling (Sarndal et al, 1992, p.129). Total angler or boat trips that fished primarily in a given area were calculated in each week by pi-expansion of the trips reported for sampled boats as follows:

$$\hat{t}_{F_a} = \frac{N_F}{n_F} \sum_{s_b} t_{F_{ab}} \quad ,$$

where $\hat{t}_{F_a}$ = the pi-estimator of the number of angler or boat trips taken primarily in area *a* on vessels in the sample frame *F*,

N_F = the total number of vessels in the sample frame,

n_F = the total number of sampled vessels for which effort data was supplied by respondents,

$t_{F_{ab}}$ = the total number of angler or boat trips reported by representatives of boat *b* as having fished primarily in area *a.*

This estimation method assumes simple random sampling of the vessels in the frame each week, and it assumes that mean reported fishing effort does not differ between non-respondent vessel representatives and respondent vessel representatives. For the purpose of comparing estimates to MRFSS estimates for different two-month periods, the VDTS estimates of effort in a particular wave were summed across weeks which included days in that wave to obtain separate effort estimates for each two-month wave.

The variance of the pi-estimator of angler fishing trips in each boat type/wave/area stratum was estimated under the assumption of simple random cluster sampling as follows:

$$\hat{V}\left(\hat{t}_{F_a}\right) = N_F^2 \frac{\left(1-\left(n_F / N_F\right)\right)}{n_F} S_{t_{F_{a_i}}}^2 \quad ,$$

where $\hat{V}\left(\hat{t}_{F_a}\right)$ = the estimated variance of the pi-estimator of the number of angler trips taken on vessels in the sample frame,

$S_{t_{F_{a_i}}}^2$ = the sample variance among boats in the total number of anglers who fished during the week.

Estimated variances of total angler fishing trip estimates were summed across strata to get estimated variances for higher level strata.

Sample data from the Observer Survey were used to estimate a ratio which could be used to adjust angler trip estimates for a given two-month sampling wave to account for trips by anglers fishing from chartered boats not included in the VDTS sample frame. The VDTS frame undercoverage correction ratio was estimated independently as follows:

$$\hat{R}_U = \frac{1}{\hat{p}_F} = \frac{n_{I_s}}{n_{IF_s}} \quad ,$$

where $\hat{R}_U$ = the estimated undercoverage correction ratio for the vessel-directory sample frame,

$\hat{p}_F$ = the estimated proportion of angler trips that fished from boats included in the vessel-directory sample frame,

n_{IF_s} = the number of angler trips intercepted by the MRFSS Intercept Survey that occurred on chartered boats included in the vessel-directory sample frame,

n_{I_s} = the total number of angler trips intercepted.

The variance of the estimated undercoverage correction ratio was estimated using the delta method as follows:

$$\hat{V}\left(\hat{R}_U\right) = \hat{V}\left(1/\hat{p}_F\right) = \frac{\left(1-\hat{p}_F\right)}{n_{I_s}\hat{p}_F^3} \quad ,$$

where $\hat{V}\left(\hat{R}_U\right)$ = the estimated variance of the undercoverage correction ratio

$\hat{V}\left(1/\hat{p}_F\right)$ = the estimated variance of the inverse of the estimated proportion $\hat{\boldsymbol{p}}_{\boldsymbol{F}}$.

The estimated correction ratio for each state/wave/mode/area stratum was applied to the appropriate estimate of angler trips from chartered boats included in the sample frame to get an estimate of total angler trips from chartered boats as follows:

$$\hat{\boldsymbol{t}}_{\boldsymbol{R}_a} = \hat{\boldsymbol{R}}_U \sum_{w} \hat{\boldsymbol{t}}_{\boldsymbol{F}_a} \ ,$$

where $\hat{\boldsymbol{t}}_{\boldsymbol{R}_a}$ = the estimator of the total number of fishing trips taken in area *a* by anglers on chartered boats,

$\hat{\boldsymbol{R}}_{\boldsymbol{U}}$ = the estimated undercoverage correction ratio,

$\hat{\boldsymbol{t}}_{\boldsymbol{F}_{a_w}}$ = the pi-estimator of the number of angler fishing trips taken on sample frame boats in area *a* during week *w*,

The accuracy of this estimator of total angler trips depends on the accuracy of the self-reported data collected by the VDTS. The variance of this estimator of total trips was estimated using Goodman's formula for the estimated variance of a product of two independent random variables (Goodman, 1960) as follows:

$$\hat{\boldsymbol{V}}\left(\hat{\boldsymbol{t}}_{\boldsymbol{R}_a}\right) = \hat{\boldsymbol{t}}_{\boldsymbol{F}_a}^2 \ \hat{\boldsymbol{V}}\left(\hat{\boldsymbol{R}}_U\right) \ + \ \hat{\boldsymbol{V}}\left(\hat{\boldsymbol{t}}_{\boldsymbol{F}_a}\right) \hat{\boldsymbol{R}}_U^2 \ - \ \hat{\boldsymbol{V}}\left(\hat{\boldsymbol{t}}_{\boldsymbol{F}_a}\right) \hat{\boldsymbol{V}}\left(\hat{\boldsymbol{R}}_U\right) \ ,$$

Under the assumption of independence, variance estimates for different boat type/wave/area strata were summed to get variance estimates at a given higher stratum level.

Estimates of the mean number of anglers per boat trip were calculated for both charter boat and headboats using the appropriate equations for estimation of a population ratio under simple random cluster sampling (Sarndal et al, 1992). The ratio of the number of anglers to the number of boat trips, R_{APBT} , was estimated for each sampling week as follows:

$$\hat{R}_{APBT_a} = \frac{\hat{t}_{AF_a}}{\hat{t}_{BF_a}},$$

where $\hat{R}_{APBT_a}$ = the estimator of the mean number of angler trips per boat trip in area *a*,

$\hat{t}_{AF_a}$ = the pi-estimator of the number of angler trips taken on sample frame boats in area *a*, and

$\hat{t}_{BF_a}$ = the pi-estimator of the number of boat trips taken on sample frame boats in area *a*.

The variance of this estimated ratio was approximated for each week as follows:

$$\hat{V}\left(\hat{R}_{APBT_a}\right) = \frac{1}{\left(\frac{\hat{t}_{BF_a}}{N_b}\right)^2}\left(\frac{\left(1-{n_b}/{N_b}\right)}{n_b}\right)\left(S_A^2 + \hat{R}_{APBT}^2 S_B^2 - 2\hat{R}_{APBT} S_{AB}\right),$$

where $\hat{V}\left(\hat{R}_{APBT_a}\right)$ = the approximated variance of the estimated ratio,

S_A^2 = the variance of the number of angler trips among sampled boats,

S_B^2 = the variance of the number of boat trips among sampled boats, and

S_{AB} = the covariance of the number of angler trips and the number of boat trips among sampled boats.

To get estimates of the mean number of angler trips per boat trip for different two-month waves, the ratio of the estimated wave totals for angler trips and boat trips was calculated. The variance of the wave level estimates was approximated using the delta method as follows:

$$\hat{V}\left(\hat{R}_{APBT_a}\right) = \frac{1}{\hat{t}_{BF}^2}\left(\hat{V}\left(\hat{t}_{AF}\right) + \frac{\hat{t}_{AF}^2 \hat{V}\left(\hat{t}_{BF}\right)}{\hat{t}_{BF}^2} - 2\frac{\hat{t}_{AF}}{\hat{t}_{BF}}\hat{C}\left(\hat{t}_{AF}, \hat{t}_{BF}\right)\right),$$

where $\hat{C}\left(\hat{t}_{AF}, \hat{t}_{BF}\right)$ = the covariance of the estimated wave totals of angler trips and boat trips.

Effort Reporting Errors

Data collected by the independent Vessel Effort Validation Survey (VEVS) was used to estimate changes needed to correct errors in the reporting of boat trips by boat representatives responding to either the VDTS or the logbook census. Reporting error corrections were estimated on the basis of differences between the reported and observed activity for a given boat on a given day. If a vessel representative reported that the boat did not take and direct observation by a VEVS sampler confirmed that the boat was in its slip, then an error correction of “0" was recorded. If the vessel representative reported that the boat did not take a trip and the VEVS observer determined that the boat was out fishing with paying passengers, then an error correction of “+1" was recorded for that day. If the vessel representative reported that the boat took a trip and the VEVS observer determined that the boat was out fishing with paying passengers, then an error correction of “0" was recorded. If the vessel representative reported that the boat took a trip and the VEVS observer determined that the boat was actually in its slip all day, then an error correction of “-1" was recorded.

For the validation of trips reported in response to the VDTS, the reported time of return of the trip and the reported number of fishing hours were used to determine whether or not a VEVS observer was present at an appropriate time to determine whether or not the trip actually occurred. The reported number of fishing hours was subtracted from the reported time of return of the trip to determine an approximate time of departure. This time was then compared with the recorded time that the boat was observed in or out of its slip by a VEVS observer. If the VEVS observer visited and observed the boat in its slip before the estimated time of departure for the reported trip, then this observation was not used in the estimates of VDTS reporting error.

Since the time of return of each trip was not included in logbook reports, estimates of logbook reporting errors assumed that VEVS observations always occurred at an appropriate time of day to verify whether or not a reported trip actually occurred. To the extent that assumption does not hold true, then some cases when boats were observed “in” at the wrong time to invalidate a reported trip would be incorrectly interpreted as evidence of an over-reporting error. Therefore, the resulting estimates of logbook reporting errors may be slightly biased toward over-reporting. In order to evaluate the possible bias caused by this assumption, estimates of VDTS reporting errors were also calculated without eliminating observations on the basis of time comparisons. This also allowed comparisons of logbook and VDTS reporting error estimates based on the same interpretations of VEVS observations.

VEVS observers occasionally made more than one dockside observation of a given boat on a given day. If more than one VEVS observation was made, only one observation was used to determine whether or not a reporting error occurred. If all observations for the day matched, then just one of those observations was used. If successive observations for the same day did not match, then priority was given to observations that confirmed that a boat was out of its slip on the given day, and among those observations priority was given to observations that confirmed that the boat was fishing with paying passengers. When more than one observation was made of a boat in its slip on a given day, then the estimated time of departure of a VDTS-reported trip for that day was compared to the time of each VEVS observation to determine if at least one occurred at an appropriate time to determine that the reported trip did not actually occur on that day.

For the purpose of estimating the total reporting error correction needed, the validation sampling of boats and days for which effort data was reported on either the VDTS or the logbook census was treated like two-stage cluster sampling, where a cluster of boats was selected in the first stage and a cluster of dates were randomly selected for dockside observations of each selected reporting boat in the second stage. The total reporting error correction needed for each wave was estimated using the formula for a pi-estimator of a population total under two-stage cluster sampling as follows:

$$\hat{t}_{E_g} = \frac{N_b}{n_b}\sum_b N_d\left(\frac{\sum_{S_d} y_{E_i}}{n_d}\right)$$

where $\hat{t}_{E_g}$ = the estimated reporting error correction for boat trips in wave *g*,

N_b = the total number of boats potentially reporting for wave *g*,

n_b = the number of boats validated in wave *g*,

y_{E_i} = the recorded error correction (+1, 0, or -1) recorded for boat *b* on day *d* of wave *g*,

N_d = the total number of days in wave *g*.

n_d = the number of days on which reported effort was validated for boat *b* in wave *g*,

This estimation method is based on the simplifying assumption that boats selected for the VDTS were randomly sampled each week by the VEVS and that all boats submitting logbook reports were randomly sampled throughout each wave. The variances of the pi-estimators of the reporting error corrections needed were estimated as follows:

$$\hat{V}\left(\hat{t}_{E_g}\right) = N_b^2\frac{\left(1-\left({}^{n_b}\!/\!_{N_b}\right)\right)}{n_b}S^2_{\hat{t}_{E_b}} + \frac{N_b}{n_b}\sum N_d^2\frac{\left(1-\left({}^{n_d}\!/\!_{N_d}\right)\right)}{n_d}S^2_{y_{E_d}},$$

where $\hat{V}\left(\hat{t}_{E_g}\right)$ = the estimated variance of the estimated total reporting error correction for wave *g*,

$S^2_{\hat{t}_{E_b}}$ = the variance of the estimated error correction among boats during wave *g*, and

$S^2_{y_{E_d}}$ = the variance of the recorded error correction among sampled days for boat *b* .

The variance of the estimated error correction among boats, $S^2_{\hat{t}_{E_b}}$, was calculated as follows:

$$S^2_{\hat{t}_{E_b}} = \frac{\sum_b\left[\hat{t}_{E_d} - \frac{\sum_b \hat{t}_{E_d}}{n_b}\right]^2}{\left(n_b - 1\right)} ,$$

where $\hat{t}_{E_d} = N_d \frac{\sum_d y_{E_i}}{n_d}$.

Due to the relatively small sample sizes for the validation survey, reporting error estimates were made at an annual level, and the estimated annual error correction was used to calculate a “reporting error correction ratio” which was used to correct estimates for each wave. This approach assumes that errors in the reporting of the number of boat trips vary among waves in proportion to the number of trips reported. Annual estimates of boat trips were corrected for estimated reporting errors by summing estimated error corrections across all six waves and adding the total correction to the annual total of boat trips reported in the logbook or estimated from the VDTS as follows:

$$\hat{t} = \hat{t}_R + \hat{t}_E = \sum_g \hat{t}_{R_g} + \sum_g \hat{t}_{E_g} ,$$

where

$\hat{t}$	=	the corrected annual estimated number of boat trips,
$\hat{t}_R$	=	the estimated annual number of boat trips based on the VDTS or logbook, and
$\hat{t}_E$	=	the estimated annual reporting error correction for the VDTS or logbook.

The appropriate reporting error correction ratio for the VDTS or the logbook was calculated as follows:

$$\hat{R}_E = \frac{\hat{t}}{\hat{t}_R} \quad .$$

The variance in the estimated reporting error adjustment ratio was estimated using the delta method as follows:

$$\hat{V}\left(\hat{R}_E\right) = \left[\frac{\left(\sum_g \hat{V}\left(\hat{t}_g\right)\right)}{\left(\sum_g \hat{t}_{R_g}\right)^2}\right] + \left[\frac{\left(\sum_g \hat{t}_g\right)^2 \left(\sum_g \hat{V}\left(\hat{t}_{R_g}\right)\right)}{\left(\sum_g \hat{t}_{R_g}\right)^4}\right] - \left[2\left(\frac{\sum_g \hat{t}_g}{\sum_g \hat{t}_{R_g}}\right) C\left(\hat{t}_g, \hat{t}_{R_g}\right)\right] ,$$

where $C\left(\hat{t}_g, \hat{t}_{R_g}\right)$ = the covariance of the estimated reporting error and the corrected estimate of total boat trips among waves.

The estimated reporting error correction ratio could be applied to the wave estimates of angler trips from the VDTS to get un-biased estimates of total angler trips from charter boats or headboats as follows:

$$\hat{t}_a = \hat{R}_E \times \hat{t}_{R_a} \quad ,$$

where $\hat{t}_a$ = the un-biased estimator of the total number of fishing trips taken in area *a* by anglers in a given wave,

$\hat{R}_E$ = the estimated reporting error correction ratio,

$\hat{t}_{R_a}$ = the pi-estimator of the number of angler fishing trips taken in area *a* during a given wave,

The logbook counts could be corrected in the same manner using the estimated reporting error correction ratio for the logbook. These estimators would be less biased since corrections would have been made for consistent reporting errors. The variance of these estimators of total trips could be estimated using Goodman's formula for the estimated variance of a product of two independent random variables (Goodman, 1960) as follows:

$$\hat{V}(\hat{t}) = \hat{t}^2 \ \hat{V}\left(\hat{R}_E\right) + \hat{V}(\hat{t}) \ \hat{R}_E^2 - \hat{V}(\hat{t}) \ \hat{V}\left(\hat{R}_E\right) \quad ,$$

Since the logbook counts were assumed to be complete reports with no variance, the calculation of the estimated variance for corrected logbook estimates would be simplified as follows:

$$\hat{V}(\hat{t}) = \hat{t}^2 \ \hat{V}\left(\hat{R}_E\right) \quad .$$

CPUE

<u>Observer Survey</u>

The traditional CPUE methods were used to calculate Intercept Survey estimates of mean catch per trip in this study. The catch data are collected by the Observer Survey to produce estimates of catch per unit effort (CPUE) that are stratified by species, catch type, and primary area of fishing. Since the area of fishing was not known in advance, intercepted angler trips were post-stratified by primary area fished for the purpose of calculating CPUE as the number of fish caught per angler trip. For each fish species, catch was separated into three different catch types – A, B1, and B2 – and independent estimates of CPUE are calculated for each catch type. Due to differences in the way the data was collected for catch types A and B, different methods are used to generate estimates of mean CPUE for the two general catch types.

Due to the difficulty that anglers often have in separating out their own individual catches when they return from a boat fishing trip, intercept survey interviewers were allowed to record the total Type A (observed) catch of a group of anglers when they interview one or more anglers in that group. When a “group catch” was recorded, the interviewer also determined and recorded the number of anglers who contributed to that catch. If the interviewed angler was able to separate his/her own Type A catch from the catch of others, then his/her individual catch was recorded and the number of contributors was recorded as “1". Interviewed anglers were always asked to report only their own individual catches of Type B fish, hence group catches were never recorded for the unobserved catch types B1 (reported removals) and B2 (reported catch released alive).

The sampling of angler trips for Type A catches was treated as simple random cluster sampling, where angler trips were randomly sampled in clusters, or groups, and catch was sampled for all trips within each sampled cluster. The mean number of Type A fish caught per angler fishing trip was estimated as the ratio of the number of fish to the number of contributors using the equation for estimation of a population ratio under simple random cluster sampling (Sarndal et al, 1992). This ratio was estimated as the ratio of the sum over all groups of the recorded Type A catches to the sum over all groups of the recorded contributors to those catches as follows:

$$\hat{R}_{FA_a} = \frac{\sum_s f_{A_a}}{\sum_s t_{A_a}}$$

where $\hat{R}_{FA_a}$ = the estimated mean number of fish caught per angler trip in area *a* based on Type A catch data alone,

$f_{A_{ai}}$ = the Type A catch recorded for group *i* of angler trips primarily fishing in area *a,*

$t_{A_{ai}}$ = the total number of angler trips contributing to the sampled Type A catch recorded for group *i* of angler trips primarily fishing in area *a*.

The estimated variance of this ratio estimator of mean Type A catch per trip was calculated as follows:

$$\hat{V}\left(\hat{R}_{FA_a}\right) = \frac{1}{\bar{t}^2_{A_a}}\left(S^2_{f_{A_a}} + \hat{R}^2_{FA_a} S^2_{t_{A_a}} - 2\hat{R}_{FA_a} S_{f_{A_a} t_{A_a}}\right)$$

where $\hat{V}\left(\hat{R}_{FA_a}\right)$ = the estimated variance of the estimated mean catch per angler trip primarily fishing in area *a* based on Type A catch,

$\bar{t}_{A_a}$ = the sample mean of the number of contributors to Type A catch among sampled groups of angler trips that primarily fished in area *a*,

$S^2_{f_{A_a}}$ = the sample variance of Type A catch among sampled groups of angler trips that primarily fished in area *a*,

$S^2_{t_{A_a}}$ = the sample variance of the number of contributors to Type A catch among sampled groups of angler trips that primarily fished in area *a*,

$s_{f_{A_a} t_{A_a}}$ = the sample covariance of Type A catch and contributors among sampled groups of angler trips that primarily fished in area *a*,

The sampling of angler trips for Type B catches (B1 and B2) was treated as simple random sampling of the catch of individual angler trips. The mean number of Type B1 or B2 fish caught per angler fishing trip was estimated as follows:

$$\hat{R}_{FB_a} = \frac{\sum_s f_{B_{a_k}}}{n_{I_a}} ,$$

where $\hat{R}_{FB_a}$ = the estimated mean number of fish of species *j* caught per angler trip based on Type B1 or B2 catch data alone,

$f_{B_{ak}}$ = the sampled Type B catch recorded for angler trip *k* that fished primarily in area *a*,

$\boldsymbol{n_{I_B}}$ = the total number of sampled angler trips for Type B catch that fished primarily in area *a*,

The estimated variance of the mean CPUE estimator based on Type B1 or B2 catch was calculated as follows:

$$\hat{V}\left(\hat{R}_{FB_a}\right) = \frac{S^2_{f_{B_a}}}{n_{I_a}} ,$$

where $\hat{V}\left(\hat{R}_{FB_a}\right)$ = the estimated variance of the estimated mean Type B catch per angler trip that fished primarily in area *a*,

$S^2_{f_{Ba}}$ = the sample variance of Type B catch per angler trip that fished primarily in area *a* .

Catch per trip estimates based on the A and B1 catch types were summed to estimate mean removals (fish landed, used for bait, or thrown back dead) per angler trip. Because estimates of mean catch per trip based on different catch types were estimated independently from the intercept survey samples, estimated variances for the A and B type catch per trip estimates were summed to estimate the variance of the estimated mean number of fish removed per trip.

CATCH ESTIMATION

Estimates Based on Observed CPUE

The estimated number of angler trips and the estimate mean catch per trip for each state/wave/mode/area stratum is used to calculate an estimate of total catch. The same methods were used in this study to estimate total catch from various estimators of total effort and mean catch per unit effort. The total catch of any given species was estimated independently for each of the three catch types. Total catch was always estimated as follows:

$$\hat{f}_{A_a} = \hat{R}_{FA_a} \times \hat{t}_a \ ,$$

where $\hat{f}_{A_a}$ = the estimated total number of fish caught in a given area,

$\hat{R}_{FA_a}$ = the estimated catch per angler trip in a given area, and

$\hat{t}_a$ = the estimated total number of angler trips in a given area.

Whenever necessary, the total angler catch of any given species in numbers of fish was estimated by summing total catch estimates across catch types. The total number of fish killed, or removed from the population, was estimated by summing the total catch estimates for the A and B1 catch types. The total number of fish caught and released alive, or returned to the population, was estimated as the total catch estimate for the B2 catch type.

The variances of the total catch estimators based on the different catch types were estimated using Goodman's formula as follows:

$$\hat{V}\left(\hat{f}_{A_a}\right) = \left(\hat{R}^2_{FA_a} \times \hat{V}\left(\hat{t}_a\right)\right) + \left(\hat{V}\left(\hat{R}_{FA_a}\right) \times \hat{t}_a^{\,2}\right) - \left(\hat{V}\left(\hat{R}_{FA_a}\right) \times \hat{V}\left(\hat{t}_a\right)\right) \ ,$$

where $\hat{V}\left(\hat{f}_{A_a}\right)$ = the estimated variance of the total catch estimator for a given area *a*.

The estimated variances were summed across catch types to get estimates of the variances of the estimators of total catch (A+B1+B2) and total removals (A + B1).